Object tracking based on incremental Bi-2DPCA learning with sparse structure

Bai, Bendu; Li, Ying; Fan, Jiulun; Price, Christopher; Shen, Qiang

Published in:
Applied Optics

DOI:
10.1364/AO.54.002897

Publication date:
2015

Citation for published version (APA):
Bai, B., Li, Y., Fan, J., Price, C., & Shen, Q. (2015). Object tracking based on incremental Bi-2DPCA learning with sparse structure. *Applied Optics*, *54*(10), 2897-2907. https://doi.org/10.1364/AO.54.002897

Object Tracking Based on Incremental Bi-2DPCA Learning with Sparse Structure

Bendu Bai,[1,*] Ying Li,[2] Jiulun Fan,[1] Chris Price,[3] and Qiang Shen[3]

[1]*School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Shaanxi, Xi'an 710121,China*

[2] *School of Computer Science, Northwestern Polytechnical University, Shaanxi, Xi'an 710072,China*

[3] *Institute of Mathematics, Physics and Computer Science, Aberystwyth University, SY23 3DB Aberystwyth, U.K*

**Corresponding author: baibendu@163.com*

Received Month X, XXXX; revised Month X, XXXX; accepted Month X, XXXX; posted Month X, XXXX (Doc. ID XXXXX); published Month X, XXXX

In this paper, we propose a novel object tracking method that can work well in challenging scenarios such as appearance changes, motion blurs, and especially partial occlusions and noise. Our method applies bilateral two-dimensional principal component analysis (Bi-2DPCA) for efficient object modeling and real-time computation requirement. An incremental Bi-2DPCA learning algorithm is proposed for characterizing the appearance changes of newly tracked objects. Also, to account for noise and occlusions, a sparse structure is introduced into our Bi-2DPCA object representation model. With this sparse structure, the appearance of an object can be represented by a linear combination of basis images and an additional noise image. The noise image, which indicates the location of noise and occlusions, can be used to effectively eliminate the influence caused by noise and occlusions and lead to a robust tracker. Instead of the reconstruction error commonly used in eigen-based tracking methods, a more accurate method is adopted for the computation of observation likelihood. The method is based on the energy distribution of coefficient matrix projected by Bi-2DPCA. Experimental results on challenging image sequences demonstrate the effectiveness of the proposed tracking method.

OCIS codes: (100.2000) Digital image processing; (100.4999) Pattern recognition, target tracking; (100.0100) Image processing; (100.4993) Pattern recognition, Baysian processors.

http://dx.doi/org/10.1364/AO.99.099999

1. Introduction

Object tracking is one of fundamental problems in the field of computer vision. It prevails in diverse applications such as intelligent surveillance, human-computer interface and vehicle navigation. Developing a robust online tracker still remains as a tough field of study, because the visual appearance of the target object may undergo large variations due to many factors such as illumination changes, pose changes, deformations and occlusions. As a result, effectively modeling the changes of the tracked object appearance plays an important role in visual tracking.

Different approaches have been proposed to address this difficult task. Among them, object tracking via online subspace learning [1]-[5] is one of attractive techniques. Instead of treating the target as a collection of low-level features [6][7], subspace representation methods provide a compact notion of the "thing" being tracked, which facilitates other vision tasks such as object recognition. In particular, Principal Component Analysis (PCA) is a popular technique for subspace learning. Li *et al.* [8] propose an incremental PCA algorithm for subspace learning. Skocaj and Leonardis [9] present a weighted incremental PCA algorithm for subspace learning. In [1], Ross *et al.* propose a generalized tracking framework based on the incremental image-as-vector PCA methods with a sample mean update. It is noted that the above tracking methods are unable to fully exploit the spatial redundancies within the image ensembles [3]. This is particularly true for those image-as-vector tracking techniques, as the local spatial information is almost lost. Consequently, the focus has been made on developing the image-as-matrix learning algorithms for effective subspace analysis. The 2-dimensional PCA (2DPCA) for image representation was developed [10], and has been adopted for object tracking [11, 12]. Based on the original image matrices, 2DPCA constructs an image covariance matrix whose eigenvectors are derived for image feature extraction. The size of the image covariance matrix of 2DPCA is much smaller in contrast with the covariance matrix of PCA. As a result, 2DPCA achieves a more efficient computation to calculate the eigenvectors. The disadvantage of 2DPCA is that 2DPCA needs more coefficients than PCA for image representation. To overcome the weakness of 2DPCA, the bilateral 2DPCA (Bi-2DPCA) [13, 14] is developed. The main idea of Bi-2DPCA is straightforward, which is to perform 2DPCA twice sequentially. The first one is in the horizontal direction and the second is in the vertical direction. At the same time, Bi-2DPCA has a solid theoretical foundation for optimal image representation mechanism in the sense of

minimal Mean Square Error (MSE). Consequently, the resulting features of Bi-2DPCA are reduced significantly, but they are still as powerful as the 2DPCA features.

It should be noted that although PCA subspace representation with online update is effective in dealing with appearance change caused by in-plane rotation, scale, illumination variation and pose change, it has also been shown that the PCA subspace based representation scheme is sensitive to partial occlusion [15].

Recently, sparse representation has been successfully applied to visual tracking [15-20]. Mei and Ling [16] were the first to develop the sparse representation technique for object tracking, in which an over-complete dictionary is constructed using target templates and trivial templates, and a l_1 regularized optimization procedure is adopted to obtain a sparse linear representation solution. Since it assumes that a target object is a linear combination of a set of target templates and trivial templates, object tracker based on sparse representation can handle noise and partial occlusions successfully. However, the l_1 tracker is computationally expensive even with further improvement [21][22], which limits its performance.

Inspired by the work above, we propose an efficient appearance model to handle the object appearance variations and partial occlusions during the tracking process. A subspace spanned by two separated projection kernels based on Bi-2DPCA and centered at the sample mean is constructed for modeling the appearance of an object. We use a new mechanism to update the subspace model online. In order to handle partial occlusions and other challenging factors, the Bi-2DPCA based object subspace model is embedded into a structure of sparse representation. Instead of the reconstruction error commonly used in eigen-based tracking methods, a more accurate method is adopted for the computation of observation likelihood. Inspired by the work in [11], our method is based on the energy distribution of the coefficient matrix projected by Bi-2DPCA. By exploiting both the capability of handling occlusions and noise of sparse representation and the power of feature extraction of Bi-2DPCA, our tracking algorithm is able to efficiently handle the high resolution image observation and the object undergoing partial occlusions, noise, and blurred. Tracking is then achieved by Bayesian inference, in which a particle filter is adopted to estimate the object state sequentially. With the tracking results in new frames, we update the subspace model adaptively. Numerous experiments demonstrate the excellence of our algorithm compared with state-of-the-art algorithms on some challenging image sequences.

The main contributions of this paper are as follows. Firstly, we employ Bi-2DPCA in a tracking framework for appearance modeling. This scheme allows us to take advantage of spatial information of raw intensity image and avoid the large size covariance matrix calculation. We propose a novel incremental Bi-2DPCA subspace learning technique which models the appearance changes of an object. Secondly, we introduce the sparse structure into our Bi-2DPCA based subspace to handle noise and partial occlusions. Finally, the time-consuming l_1 regularized optimization procedure is replaced by an iterative algorithm to achieve a real-time implementation.

The remaining part of this paper is organized as follows. Section 2 reviews the basic conceptions of Bi-2DPCA and sparse representation based object tracking. Our appearance model for visual tracking is described in Section 3. Section 4 presents the proposed tracking algorithm and the incremental Bi-2DPCA learning algorithm. Section 5 gives our experiments and performance evaluation of our tracker and we conclude this paper in Section 6.

2. Related work

In this section, we will review the most relevant theories and algorithms associated with our tracking method.

A. Bi-2DPCA and its properties

We review 2DPCA first, as the main idea of Bi-2DPCA is to perform 2DPCA in the horizontal and vertical direction sequentially. Given a 2D image $\mathbf{A}$, expressed as a $m \times n$ matrix, the aim of 2DPCA is to project image $\mathbf{A}$ onto a unitary column vector $\mathbf{X}$ by the following linear transformation

$$\mathbf{Y} = \mathbf{AX} \tag{1}$$

where $\mathbf{Y}$ is a m-dimensional projected vector. The criterion used to determine a good projection vector $\mathbf{X}$ is to maximize the total scatter of the projected vectors. We recall that the total scatter measures the discriminatory power of the projection vector $\mathbf{X}$. The total scatter of the projected vectors can be characterized by the trace of the covariance matrix of the projected vectors. As a consequence, we adopt the following criterion

$$J(\mathbf{X}) = tr(\mathbf{S}_\mathbf{Y}) \tag{2}$$

where $\mathbf{S}_\mathbf{Y}$ is the covariance matrix of the projected vectors and $tr(\mathbf{S}_\mathbf{Y})$ denotes the trace of $\mathbf{S}_\mathbf{Y}$, which takes the form

$$tr(\mathbf{S}_\mathbf{Y}) = \mathbf{X}^T[E(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})]\mathbf{X} \tag{3}$$

Then a $n \times n$ non-negative definite matrix is defined as

$$\mathbf{G}_t = E(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A}) \tag{4}$$

The symmetric matrix $\mathbf{G}_t$ is called the image covariance scatter matrix. It can be evaluated by the training samples of the random matrix $\mathbf{A}$. Suppose that there are N training image samples $\{\mathbf{A}_i\}_{i=1}^N$, and $\overline{\mathbf{A}}$ denotes their mean. Then, $\mathbf{G}_t$ can be computed approximately by

$$\mathbf{G}_t = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{A}_i - \overline{\mathbf{A}})^T(\mathbf{A}_i - \overline{\mathbf{A}}) \tag{5}$$

Then the criterion in (2) can be rewritten by

$$J(\mathbf{X}) = \mathbf{X}^T\mathbf{G}_t\mathbf{X} \tag{6}$$

Based on criterion in (6), the optimal projection axes $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_q$ are chosen as the orthogonal eigenvectors of $\mathbf{G}_t$ corresponding to the first q $(q < n)$ largest

eigenvalues. Let $\mathbf{R} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_q]$ denote the right projection kernel. The projected matrix of $\mathbf{A}_t$ can be expressed by

$$\mathbf{B}_i = (\mathbf{A}_i - \overline{\mathbf{A}})\mathbf{R} \tag{7}$$

After the projection of image patches through the right projection kernel, the projected matrix eliminates the correlations between image columns and compresses the image energy optimally in the horizontal direction. As 2DPCA directly handles image patches by a 2D matrix, it has an advantage in preserving spatial information, and in calculating the covariance matrix. However the compression rate of 2DPCA is far lower than PCA and more coefficients are required for the image representation. This leads to a large storage and significant computation requirement.

Bi-2DPCA can overcome the weakness of 2DPCA. We construct the covariance matrix $\mathbf{F}_t$ based on $\mathbf{B}_i^T$ and it can be evaluated by

$$\mathbf{F}_t = \frac{1}{N}\sum_{i=1}^{N} \mathbf{B}_i^T \mathbf{B}_i \tag{8}$$

Let $\mathbf{L} = [\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_p]$ $(p < m)$ be the left projection kernel, where its columns are the orthogonal eigenvectors of $\mathbf{F}_t$ corresponding to the p largest eigenvalues $\lambda_L^1 \geq \lambda_L^2 \geq ... \geq \lambda_L^p$. The coefficient matrix based on Bi-2DPCA of a specified image patch $\mathbf{A}_i$ is formulated as

$$\mathbf{C}_i = \mathbf{L}^T(\mathbf{A}_i - \overline{\mathbf{A}})\mathbf{R} \tag{9}$$

The resulting coefficient matrix $\mathbf{C}_i$ is a $p \times q$ matrix, whose size is much smaller than the projected matrix and the original image patch since p and q are usually selected much smaller than m and n. The main process of Bi-2DPCA is illustrated in Fig.1.

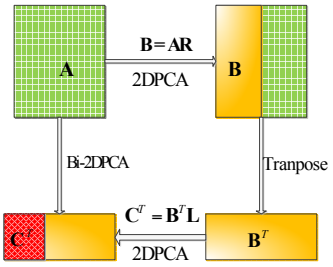


Fig.1 Illustration of Bi-2DPCA

B. Object tracking via sparse representation

Sparse representation has been successfully applied to object tracking by Mei and Ling [16]. They assume that an object observation can be linearly represented by a basis library that consists of a target template set and occlusion template set. They cast the tracking problem as

$$\mathbf{y} = \mathbf{Tz} + \mathbf{e} = \begin{bmatrix}\mathbf{T} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{z} & \mathbf{e}\end{bmatrix}^T = \mathbf{Bc} \tag{10}$$

where $\mathbf{y}$ is an observation vector, $\mathbf{T}$ denotes a matrix of target templates, $\mathbf{z}$ represents the corresponding coefficients, and $\mathbf{e}$ is the error vector, a fraction of whose entries are nonzero when y is the observation of a true object. The error caused by noise and occlusions typically corrupts a fraction of the image pixels. We get a sparse solution of (10)

$$\min_c \ \|\mathbf{Bc} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{c}\|_1 \tag{11}$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ denote the l_2 and l_1 norms respectively, λ is a parameter that balances the tradeoff between the reconstruction error and the sparsity. This formulation leads to the well-known l_1 tracker which is time-consuming due to l_1 minimization problem.

Because the sparsity of $\mathbf{Z}$ is not required and the target templates of $\mathbf{T}$ are mutually correlated, each object observation can be sparsely approximated by the learned eigenbases and other noise bases. So, Wang *et al.* [15] cast object tracking as the following optimization problem

$$\min_{\mathbf{z},\mathbf{e}} \ \|\mathbf{y} - \mathbf{Uz} - \mathbf{e}\|_2^2 + \lambda\|\mathbf{e}\|_1 \tag{12}$$

where $\mathbf{U}$ is an orthogonal matrix of column PCA basis vectors. For (12), it can handle partial occlusions effectively, and compared with the l_1 tracker, it has less computational complexity by exploiting the subspace representation power and a new iterative scheme [15].

For (11), referred to as the l_1 tracker, has two drawbacks. One is the extensive computational cost brought by l_1 minimization as already discussed. Another is the expressiveness of the object appearance is limited as the object appearance can only be represented by the subspace spanned by the target templates directly cropped from the images. This makes it difficult to handle significant appearance changes and fast motion. Different from the l_1 tracker, the tracking method based on (12) represent an object with orthogonal basis vectors extracted from a collection of tracked objects, which depict the statistical model of samples. This method achieves a relatively good performance. However, PCA is inefficient at handling the high-dimensional feature vectors. Below we will introduce a novel object appearance model based on Bi-2DPCA that can handle the problems which PCA does not.

3. Object appearance modeling

We present the detail of the proposed object appearance model here. In the proposed appearance model, a subspace spanned by two separated projection kernels based on Bi-2DPCA and centered at the sample mean, is constructed for modeling the appearance of the tracked objects. Then the Bi-2DPCA based subspace model is embedded into a sparse structure to account for noise and occlusions.

A. Subspace model based on Bi-2DPCA

Suppose we have a collection of observations $\{\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_t\}$ which each element $\mathbf{A}_i$ is a $m \times n$ matrix normalized from the object image region of the ith frame. We obtain the covariance matrices $\mathbf{G}_t$ and $\mathbf{F}_t$ based on (5) and (8). Suppose the right projection kernel $\mathbf{R}_t = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_q]$ is a $n \times q$ matrix that consists of the orthogonal eigenvectors of $\mathbf{G}_t$ corresponding to the q largest eigenvalues $\lambda_R^1 \geq \lambda_R^2 \geq ... \geq \lambda_R^q$; the left projection kernel $\mathbf{L}_t = [\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_p]$ is a $m \times p$ matrix that consists of the orthogonal eigenvectors of $\mathbf{F}_t$ corresponding to the p largest eigenvalues $\lambda_L^1 \geq \lambda_L^2 \geq ... \geq \lambda_L^p$. Let $\mathbf{\Sigma}_R^t = diag(\lambda_R^1, \lambda_R^2, ..., \lambda_R^q)$ and $\mathbf{\Sigma}_L^t = diag(\lambda_L^1, \lambda_L^2, ..., \lambda_L^p)$ .The average image of all image patches at time t is denoted by $\overline{\mathbf{A}}_t$. Next, we describe the object subspace model. The subspace is spanned by $\mathbf{R}_t$ and $\mathbf{L}_t$, centered at $\overline{\mathbf{A}}_t$. $\mathbf{\Sigma}_R^t$ and $\mathbf{\Sigma}_L^t$ evaluate the energy distribution of all training samples along each projection vector.

Compared with the conventional PCA based subspace model, the Bi-2DPCA based subspace has the following advantages. Firstly, the Bi-2DPCA based subspace does not face the massive computation caused by the high-dimensional feature vectors, because Bi-2DPCA directly uses the original object image matrix. It is easy to calculate the eigenvectors and eigenvalues of its covariance matrices. In the following, we will see that it needs less computation to recursively update the subspace for charactering the appearance changes of the tracked object. Secondly, Bi-2DPCA has a solid theoretical foundation for optimal image representation mechanism in the sense of minimal MSE. Furthermore, the efficiency of Bi-2DPCA does not depend on the distribution of data while PCA does. Thus, the Bi-2DPCA based space is particularly suited for object appearance characterization.

B. Bi-2DPCA based subspace with sparse structure

The original image can be reconstructed by

$$\mathbf{A} = \overline{\mathbf{A}} + \mathbf{L}\mathbf{C}\mathbf{R}^T \tag{13}$$

Denote $\mathbf{C} = (c_{ij})_{p \times q}$, (13) can be rewritten by

$$\mathbf{A} = \overline{\mathbf{A}} + \sum_{i=1}^{p}\sum_{i=1}^{q} c_{ij}\mathbf{W}_i\mathbf{X}_j^T \tag{14}$$

Denote $\mathbf{\psi}_{ij} = \mathbf{W}_i\mathbf{X}_j^T$. Obviously, $\mathbf{\psi}_{ij}$ is a rank-1 matrix, which is the same size of the original image patch $\mathbf{A}$ and is called the basis image [14] corresponding to the basis vectors of PCA. Any image can be approximately reconstructed by adding up the weighted basis images and the mean image. In order to handle occlusions and noise, we add an error matrix $\mathbf{O} = (o_{ij})_{m \times n}$ to the right side of (14) and ignore the mean for the moment

$$\mathbf{A} = \sum_{i=1}^{p}\sum_{j=1}^{q} c_{ij}\mathbf{W}_i\mathbf{X}_j^T + \sum_{i=1}^{m}\sum_{j=1}^{n} o_{ij}\mathbf{e}_i\mathbf{u}_j^T \tag{15}$$

where $\mathbf{e}_i \in R^m$ and $\mathbf{u}_j \in R^n$ are both column vectors with only one nonzero entry 1 located at i and j respectively. Similar to the l_1 tracker, the error term $\mathbf{O}$ is caused by occlusions and noise, which typically corrupts a fraction of the image pixels. Therefore, for a true object, there is only a limited number of nonzero entries in the error term $\mathbf{O}$. So, we want to have a sparse solution to (15). Thus,

$$\min_{C,O} \left\| \mathbf{A} - \mathbf{L}\mathbf{C}\mathbf{R}^T - \mathbf{O} \right\|_F + \lambda \left\| \mathbf{O} \right\|_1 \tag{16}$$

where λ is a parameter that balances the tradeoff between the reconstruction error and the sparsity, $\|\cdot\|_F$ is F norm of a matrix, $\|\cdot\|_1$ means summing the absolute values of nonzero entries of a matrix. Fig.2 gives the object representation schemes of the three formulations defined in Eq. (11), (12) and (16).

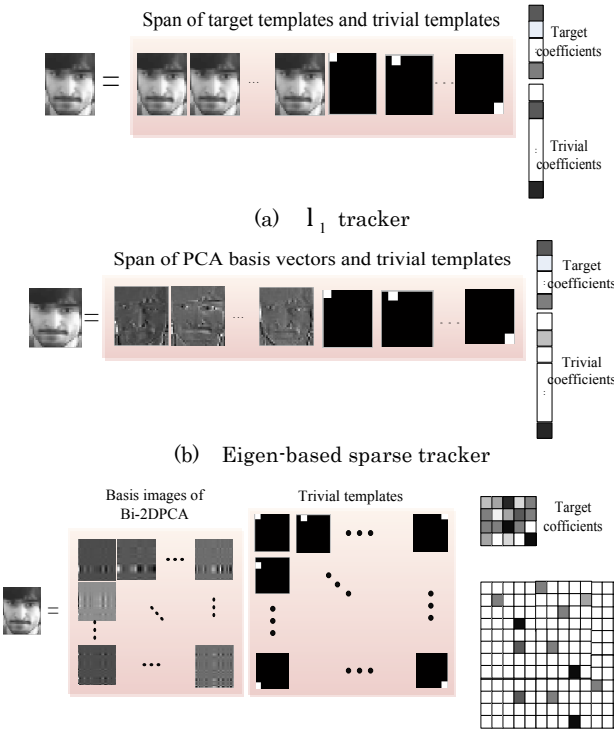


Fig.2. Three object representation schemes

Our implementation solves the optimization problem in (16) via an iterative algorithm, which is an extension of the algorithm in [15] into 2D matrix space. We demonstrate its feasibility through many experiments. The algorithm is shown in Algorithm 1.

Because the Bi-2DPCA based subspace representation model aims to adapt the appearance variations with a low-dimensional subspace for dimensionality reduction and real-time calculation, it is efficient and effective for handling the situation where the object undergoes pose,

scale and illumination changes. This representation model, however usually fails in the presence of occlusions and noise. Sparse representation based appearance model treats an object as a linear combination of target templates and an additional noise item which is explicitly designed for modeling the occlusions and noise. Therefore the sparse structure is introduced into the Bi-2DPCA object representation model to take into account both advantages of two models.

Algorithm 1: The iterative algorithm for Eq. (16)

Input: an observation image $\mathbf{A}$, projection kernels $\mathbf{L},\mathbf{R}$, and a constant λ.

1: Initialize $\mathbf{O}$ be a zero matrix;

2: Iterate

3: Obtain $\mathbf{C} = \mathbf{L}^T(\mathbf{A}-\mathbf{O})\mathbf{R}$;

4: Obtain $\mathbf{O} = \mathbf{S}_\lambda(\mathbf{A}-\mathbf{LCR}^T)$, where $S_\lambda(x) = \mathrm{sgn}(x)\cdot(|x|-\lambda)$;

5: Until $\Delta J(\mathbf{C},\mathbf{O}) < \gamma$ or a certain number iterations arrive, where $J(\mathbf{C},\mathbf{O}) = \left\|\mathbf{A}-\mathbf{LCR}^T-\mathbf{O}\right\|_F + \lambda\|\mathbf{O}\|_1$ and γ is a small positive number.

Output: the coefficient matrix $\mathbf{C}$, the error matrix $\mathbf{O}$.

4. Proposed tracking algorithm

Object tracking can be cast as a Bayesian inference process for estimating the unknown state $\mathbf{x}_t$. The posterior distribution $p(\mathbf{x}_t \mid \mathbf{y}_{1:t})$ over $\mathbf{x}_t$ is recursively updated given all observations $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t\}$ up to time t.

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t \mid \mathbf{x}_t)\int p(\mathbf{x}_t \mid \mathbf{x}_{t-1})p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (17)$$

where $p(\mathbf{y}_t \mid \mathbf{x}_t)$ is the observation model depicting the likelihood of observing $\mathbf{y}_t$ at state $\mathbf{x}_t$, and the dynamic model $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ defines motion randomness between two consecutive states. In the particle filter, the posterior distribution $p(\mathbf{x}_t \mid \mathbf{y}_{1:t})$ is approximated by discrete random measures defined by particles $\{\mathbf{x}_t^i\}_{i=1}^N$ and weights $\{\mathbf{w}_t^i\}_{i=1}^N$ assigned to the particles. The optimal state $\mathbf{x}^*$ of the tracked object, given all the observations up to t-frame, is obtained by the MAP estimation over N samples at time t

$$\mathbf{x}^* = \arg\max_{\mathbf{x}_t^i} \mathbf{w}_t^i \ , \ i=1,2,...,N \quad (18)$$

where $\mathbf{w}_t^i = \mathbf{w}_{t-1}^i p(\mathbf{y}_t \mid \mathbf{x}_t^i)$.

A. Dynamic model

In the tracking framework, we apply an affine image warping to model the target motion between two consecutive frames. The six parameters of the affine transform are used to model $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ of a tracked target. Let $\mathbf{x}_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)$, where x_t, y_t, θ_t, s_t, α_t, and ϕ_t denote the x, y translations, the rotation angle, the scale, the aspect ratio, and the skew direction at time t, respectively. Generally, the dynamic model is formulated by a Gaussian distribution as follows:

$$p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = N(\mathbf{x}_t; \mathbf{x}_{t-1}, \boldsymbol{\Phi}) \quad (19)$$

where $\boldsymbol{\Phi}$ is a diagonal covariance matrix whose elements depict how much we expect the object might move from one frame to the next. The value of $\boldsymbol{\Phi}$'s diagonal elements affect the accuracy of our tracking process. With smaller values in the diagonal covariance matrix $\boldsymbol{\Phi}$, we may lose the object, and larger values may need more particles. It is our task to find a balance.

B. Observation model

The object appearance can be modeled by a Bi-2DPCA based subspace, spanned by the separate projection kernels $\mathbf{R}_t$ and $\mathbf{L}_t$, centered at $\overline{\mathbf{A}_t}$, embedded into a sparse structure. The observation model is used to evaluate the probability of a sample being generated from our appearance model. Given an image patch $\mathbf{A}_t^i$ predicated by the candidate target state $\mathbf{x}_t^i$ in the current frame, the probability of $\mathbf{A}_t^i$ is generated from our appearance model can be expressed by the reconstruction error as following

$$p(\mathbf{y}_t \mid \mathbf{x}_t^i) = \exp\left[-\frac{1}{2}\left\|\mathbf{A}_t^i - \overline{\mathbf{A}_t} - \mathbf{L}_t\mathbf{C}_t^i\mathbf{R}_t^T\right\|_F\right] \quad (20)$$

where $\mathbf{C}_t^i$ is the coefficient matrix of $\mathbf{A}_t^i$, and $\mathbf{C}_t^i = \mathbf{L}_t^T\mathbf{A}_t^i\mathbf{R}_t$.

However the reconstruction error has the drawback that it does not consider the distance from the center of subspace to the projected point, and ignores the relationship between the projection axes, which causes inaccuracies when applying the MAP for state estimation. The distance should instead be computed according to the energy distribution of each of the projection axes. Here we propose a new computation scheme that can sovle the above problem, expressed as

$$\begin{aligned} p(\mathbf{y}_t \mid \mathbf{x}_t^i) \propto & \exp\left[-\frac{1}{2\sigma^2}\left\|\mathbf{A}_t^i - \overline{\mathbf{A}_t} - \mathbf{L}_t\mathbf{C}_t^i\mathbf{R}_t^T\right\|_F\right] \\ & \times \exp\left[-\frac{1}{2}\sum_{k=1}^{p}\sum_{l=1}^{q}\frac{\mathbf{C}_t^i(k,l)^2}{\lambda_L^k + \lambda_R^l}\right] \end{aligned} \quad (21)$$

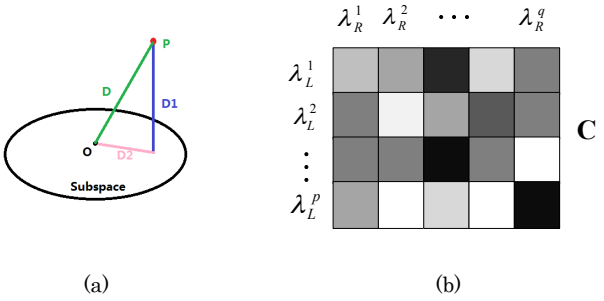
where $\sigma^2 = \frac{1}{m-p}\sum_{i=p+1}^{m}\lambda_L^i + \frac{1}{n-q}\sum_{i=q+1}^{n}\lambda_R^i$, $\lambda_L^1 \geq \lambda_L^2 \geq ... \geq \lambda_L^m$ and $\lambda_R^1 \geq \lambda_R^2 \geq ... \geq \lambda_R^n$ are eigenvalues of the covariance matrix $\mathbf{G}_t$ and $\mathbf{F}_t$ respectively. $\mathbf{C}_t^i(k,l)$ denotes the k-th row and the l-th column element of the coefficient matrix $\mathbf{C}_t^i$.

Fig. 3(a) and 3(b) illustrate Eq. (21) intuitively. As shown in Fig. 3(a), the region inside the black circle indicates the subspace generated by Bi-2DPCA, where O is the subspace center, P is the location of the candidate target, D is the overall distance from P to O, and Dl and

$D2$ are two orthogonal components of D respectively. In real tracking applications, it is necessary for a subspace-based tracker to evaluate the likelihood between the candidate target and the learned subspace. Commonly, this likelihood is determined by the reconstruction error norm, actually which is the component $D1$ in Fig. 3(a). Obviously, another orthogonal component $D2$ in the subspace should not be negligible. Therefore, we use D instead of $D1$ for observation modeling. In this paper, we calculate the component $D2$ by

$$D2 = \sum_{k=1}^{p}\sum_{l=1}^{q}\frac{\mathbf{C}_t^i(k,l)^2}{\lambda_L^k + \lambda_R^l} \tag{22}$$

On the other hand, Fig. 3(b) intuitively shows the energy distribution of the coefficient matrix projected by projection kernels of Bi-2DPCA. Given an image patch $\mathbf{A}$, we can obtain the coefficient matrix $\mathbf{C}$ via a complete Bi-2DPCA transform, whose entry corresponds to a small block in Fig. 3(b). The energy of the image is reassigned on $\mathbf{C}$. The i-th column of $\mathbf{C}$ is the projection of the image $\mathbf{A}$ on the i-th column vector of the right projection matrix $\mathbf{R}$ corresponding to the eigenvalue λ_R^i, which can be thought of as its allocated energy. Similarly, the j-th row of $\mathbf{C}$ indicates the allocated energy of the j-th row vector of the left projection matrix $\mathbf{L}$ corresponding to the eigenvalue λ_L^j. As a result, the contribution of the entry of $\mathbf{C}$ for computing the distance D2 depends on its allocated energy. Thus, (22) can describe the distance more accurately for the tracking problem, which is similar to the Mahalanobis distance.



(a) (b)

Fig. 3 (a) Decomposition of distance. (b) The energy distribution of the coefficient matrix projected by projection kernels of Bi-2DPCA.

C. Incremental Bi-2DPCA learning

During tracking, the object appearance may change due to many factors such as illumination changes, pose changes, deformations and occlusions. Therefore, it is inaccurate to maintain a fixed object appearance model. Adapting the appearance models online to reflect these changes is crucial to keeping the tracker robust. Our object appearance model is a combination of the Bi-2DPCA based subspace and the object sparse representation. We only need to update the subspace, because it is the only part that corresponds to the changes of the object appearance. Let $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_t\}$ be the original observation set and let $\mathbf{B} = \{\mathbf{A}_{t+1}, \mathbf{A}_{t+2}, ..., \mathbf{A}_{t+M}\}$ be the newly added observation set, then the total observation set at time $t+M$ can be written as $\mathbf{C} = \{\mathbf{A}, \mathbf{B}\} = \{\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_{t+M}\}$. Denote $\overline{\mathbf{A}_t}$, $\overline{\mathbf{A}_M}$, $\overline{\mathbf{A}_{t+M}}$ as the corresponding average images of sets $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$. The related covariance matrices can be computed as follows

$$\mathbf{G}_t = \frac{1}{t}\sum_{i=1}^{t}(\mathbf{A}_i - \overline{\mathbf{A}_t})^T(\mathbf{A}_i - \overline{\mathbf{A}_t}) \tag{23}$$

$$\mathbf{G}_M = \frac{1}{M}\sum_{i=t+1}^{t+M}(\mathbf{A}_i - \overline{\mathbf{A}_M})^T(\mathbf{A}_i - \overline{\mathbf{A}_M}) \tag{24}$$

$$\mathbf{F}_t = \frac{1}{t}\sum_{i=1}^{t}(\mathbf{A}_i - \overline{\mathbf{A}_t})\mathbf{R}_{t+M}\mathbf{R}_{t+M}^T(\mathbf{A}_i - \overline{\mathbf{A}_t})^T \tag{25}$$

$$\mathbf{F}_M = \frac{1}{M}\sum_{i=t+1}^{t+M}(\mathbf{A}_i - \overline{\mathbf{A}_M})\mathbf{R}_{t+M}\mathbf{R}_{t+M}^T(\mathbf{A}_i - \overline{\mathbf{A}_M})^T \tag{26}$$

$$\mathbf{G}_{t+M} = \frac{1}{t+M}\sum_{i=1}^{t+M}(\mathbf{A}_i - \overline{\mathbf{A}_{t+M}})^T(\mathbf{A}_i - \overline{\mathbf{A}_{t+M}}) \tag{27}$$

$$\mathbf{F}_{t+M} = \frac{1}{t+M}\sum_{i=1}^{t+M}(\mathbf{A}_i - \overline{\mathbf{A}_{t+M}})\mathbf{R}_{t+M}\mathbf{R}_{t+M}^T(\mathbf{A}_i - \overline{\mathbf{A}_{t+M}})^T \tag{28}$$

where $\mathbf{R}_t$, $\mathbf{R}_{t+M}$ are matrices constructed by the Eigen-vectors of covariance matrices $\mathbf{G}_t$ and $\mathbf{G}_{t+M}$ corresponding to the q largest eigenvalues. Then we have the following results

$$\mathbf{G}_{t+M} = \frac{1}{t+M}\left[t\mathbf{G}_t + M\mathbf{G}_M + \frac{t\cdot M}{t+M}(\overline{\mathbf{A}_M} - \overline{\mathbf{A}_t})^T(\overline{\mathbf{A}_M} - \overline{\mathbf{A}_t})\right] \tag{29}$$

$$\mathbf{F}_{t+M} = \frac{1}{t+M}\left[t\mathbf{F}_t + M\mathbf{F}_M + \frac{t\cdot M}{t+M}(\overline{\mathbf{A}_M} - \overline{\mathbf{A}_t})\mathbf{R}_{t+M}\mathbf{R}_{t+M}^T(\overline{\mathbf{A}_M} - \overline{\mathbf{A}_t})^T\right] \tag{30}$$

However $\mathbf{F}_t$ in (30) is computationally infeasible as the right projection matrix is changed from $\mathbf{R}_t$ to $\mathbf{R}_{t+M}$, it is impossible to store all the observation images and it is also time-consuming to compute $\mathbf{F}_t$. So we use

$$\tilde{\mathbf{F}}_t = \frac{1}{t}\sum_{i=1}^{t}(\mathbf{A}_i - \overline{\mathbf{A}_t})\mathbf{R}_t\mathbf{R}_t^T(\mathbf{A}_i - \overline{\mathbf{A}_t})^T \tag{31}$$

instead of $\mathbf{F}_t$ and obtain

$$\mathbf{F}_{t+M} \approx \frac{1}{t+M}\left[t\tilde{\mathbf{F}}_t + M\mathbf{F}_M + \frac{t\cdot M}{t+M}(\overline{\mathbf{A}_M} - \overline{\mathbf{A}_t})\mathbf{R}_{t+M}\mathbf{R}_{t+M}^T(\overline{\mathbf{A}_M} - \overline{\mathbf{A}_t})^T\right] \tag{32}$$

Lots of experiments demonstrate its feasibility. The major computation of (29) and (32) is to evaluate $\mathbf{G}_M$ and $\mathbf{F}_M$. Because M, that is the batch size for our appearance model update, is often a small value, the time and space consumption of incremental Bi-2DPCA learning method is reduced greatly. Algorithm 2 presents its overall learning process.

Algorithm 2. Incremental Bi-2DPCA Learning

Input: the mean image $\mathbf{A}_t$, projection kernels $\mathbf{L}_t$, $\mathbf{R}_t$, covariance matrices $\mathbf{G}_t$, $\mathbf{F}_t$, the newly added observation set $\mathbf{B}$ and its mean $\overline{\mathbf{A}_M}$;

1. Compute the mean image $\overline{\mathbf{A}_{t+M}} = \frac{1}{t+M}(t \cdot \overline{\mathbf{A}_t} + M \cdot \overline{\mathbf{A}_M})$.
2. Obtain the covariance matrix $\mathbf{G}_{t+M}$ by (29).
3. Compute the EVD (Eigen Value Decomposition) of $\mathbf{G}_{t+M}$: $\mathbf{G}_{t+M} = \mathbf{R}\mathbf{\Sigma}_R\mathbf{R}^T$. $\mathbf{R}_{t+M}$ consists of the columns of $\mathbf{R}$ corresponding the q largest eigenvalues of $\mathbf{G}_{t+M}$, and $\mathbf{\Sigma}_R^{t+M}$ is a diagonal matrix whose elements are the q largest eigenvalues of $\mathbf{G}_{t+M}$.
4. Obtain the covariance matrix $\mathbf{F}_{t+M}$ by (32).
5. Compute the EVD of $\mathbf{F}_{t+M}$: $\mathbf{F}_{t+M} = \mathbf{L}\mathbf{\Sigma}_L\mathbf{L}^T$, $\mathbf{L}_{t+M}$ consists of the columns of $\mathbf{L}$ corresponding the p largest eigenvalues of $\mathbf{F}_{t+M}$, and $\mathbf{\Sigma}_L^{t+M}$ is a diagonal matrix whose elements are the p largest eigenvalues of $\mathbf{F}_{t+M}$.

Output: the mean image $\overline{\mathbf{A}_{t+M}}$, projection kernels $\mathbf{L}_{t+M}$, $\mathbf{R}_{t+M}$ and its corresponding eigenvalues $\mathbf{\Sigma}_L^{t+M}$, $\mathbf{\Sigma}_R^{t+M}$, covariance matrices $\mathbf{G}_{t+M}$, $\mathbf{F}_{t+M}$.

5.Experimental results

To demonstrate performance of the proposed tracking algorithm, we tested it on 8 publicly available image sequences collected from the public dataset [15,23]. These image sequences involve different situations that an object may experience such as illumination changes, deformations, partial occlusions, fast movement, facial expression changes and large pose variations. The resolution, length and challenges in each sequence are listed in Table 1. We also compare our tracker with four other state-of-the-art tracking methods using the source codes provided by the authors, including FRAG [24], l_1-APG [22], CT [25], and IVT [1].

Table 1 The used test sequences and their challenges

Sequences	Frame size	Frame length	Challenges
Occlusion1	352×288	889	Partial occlusion
Occlusion2	320×240	819	Partial occlusion, in-plane rotation, out-plane rotation
Car4	360×240	450	Illumination variation, scale change
Car11	320×240	393	Illumination variation, background clutter, scale change
David_indoor	320×240	462	Illumination variation, scale change, out-plane rotation
Caviar1	384×288	382	Partial occlusion, scale change
Lemming	640×480	1336	out-plane rotation, scale change, occlusion, background clutter
Cliffbar	320×240	471	in-plane rotation, scale change, background clutter, abrupt motion

Without loss of generality, we locate the target object manually in the first frame using a rectangle, and transform all image sequences to the gray scale. The object region used in our experiments is normalized to a 32×32 patch. In the previous 8 frame images, we use the classical particle filter algorithm to obtain a target object sample set. Then we use the observations of these target image patches to initialize our Bi-2DPCA based subspace. The subspace is updated every 5 frames. The number of columns of the right projection matrix $\mathbf{R}_t$ is set to 6, the same as the left projection matrix $\mathbf{L}_t$. The number of particles is set to 600.

In the *Occlusion1* sequence, a woman occludes her face with a book frequently. The main difficulty of the tracking is occlusion issue in this video sequence. The tracked object is even covered 50%-80% in some video frames, such as the 534th frame and the 622th frame. As shown in Fig. 4(a), the FRAG, l_1-APG and the proposed algorithm perform better than the CT tracker. The FRAG method uses the part-based representation with histograms to handle occlusion. The proposed and l_1-APG trackers handle partial occlusions via the sparse representation with trivial templates.

In the *occlusion2* sequence, the proposed tracker performs best although the face is heavily occluded with in-plane rotations such as the 473th and 716th frames as shown in Fig. 4(b). In these frames, the FRAG algorithm performs poorly since it does not handle appearance change caused by pose and occlusion.

Fig. 4(c) shows the tracking results of the *Car4* sequence, which is captured in an open road scenario. There is a drastic lighting change when the vehicle goes underneath the overpass or the trees. The l_1-APG, IVT and the proposed algorithm track the target quite well whereas the other methods drift away when drastic illumination or scale variation occurs.

In the *Car11* sequence, a car is driven into a very dark environment, while being videotaped from another moving car. The Fig. 4(d) shows the tracking results. Between the 10th frame and the 172th frame, although the appearance of the target does not change a lot, the FRAG and CT trackers lose the target due to the low contrast between the foreground and the background and illumination changes. The car starts to turn from the 280th frame, the l_1-APG algorithm loses the target gradually and the IVT algorithm gets drift finally. The proposed tracker performs well in the whole sequence.

In the *David_indoor* sequence, shown in Fig. 4(e), the appearance of the man changes significantly when he walks from a dark room into areas with spot light. In addition, appearance change caused by scale and pose as well as camera motion pose great challenges. The IVT and proposed methods perform quite robustly to appearance and illumination changes in this case, while the other trackers do not adapt to scale or in-plane rotation.

Fig. 4(f) shows the tracking results of the *Caviar1* sequence. The main difficulty of tracking for this video lies

in heavy occlusions and scale changes. In addition, there are numerous objects with similar appearance (shape and color) to the target. The IVT tracker often leads to drifts due to the simple update scheme without dealing with occluded regions. In contrast, the FRAG and proposed methods achieve stable performance in the entire sequences in spite of heavy occlusions and large scale changes.

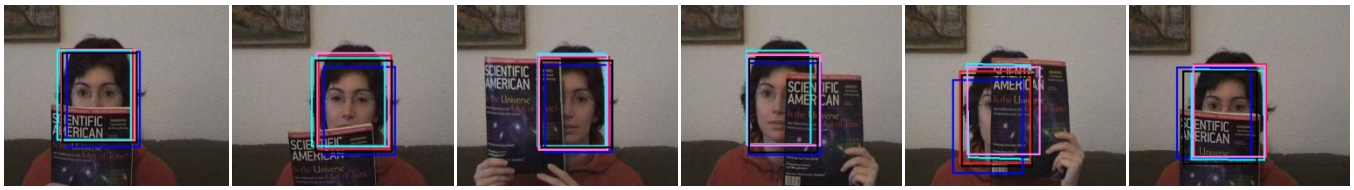
Fig. 4(g) and Fig. 4(h) show tacking results from two challenging sequences with complex background. For the *Lemming* sequence shown in Fig.10, the object undergoes change of scale and pose, as well as heavy occlusion in cluttered background. The object in the *Cliffbar* sequence undergoes scale change, in-plane rotation, and abrupt motion in cluttered background. In addition, the target and the surrounding region have similar texture. The l_1-APG, CT and FRAG algorithms perform poorly since the surrounding background is similar to the target object. The IVT method is able to track the targets in some frames but fails after abrupt motion occurs. However, our tracker successfully keeps track of the target objects with small errors in these two sequences.

To quantitatively compare the performance of the tracking methods, we compute the center location error (CLE) which is defined as the average Euclidean distance between the center locations of the tracked targets and the ground truths, as well as the average CLE over all the frames of one sequence. The average CLE of each tracker on all sequences are shown in Table 2. Fig. 5 shows the CLE of each tracker over time on each sequence. It can be seen that our method achieves lowest tracking errors in almost all the sequences.

Our tracking method is implemented in MATLAB which runs on a PC with Intel Core 3.00GHz CPU. Without code optimization, our tracking algorithm can achieve about 5-6 frames per second (fps) for the evaluated sequences. Its running speeds could be increased remarkably by using C/C++ and code optimization. Actually, if we conservatively predict a 5-time growth, our tracking method will become qualified for real-time applications.

Table 2. Average CLE (in pixel) for five trackers

Sequence	FRAG	l_1-APG	CT	IVT	Ours
Occlusion1	5.6	6.1	27.0	9.2	**2.4**
Occlusion2	15.5	8.9	12.5	10.2	**4.7**
Car4	179.8	4.1	153.7	2.9	**2.6**
Car11	63.9	3.1	76.8	**2.1**	**2.1**
David_indoor	76.7	16.1	25.6	3.6	**2.3**
Caviar1	5.7	8.9	17.1	45.1	**5.6**
Lemming	149.1	184.5	135.9	93.4	**8.5**
Cliffbar	48.7	83.6	21.9	24.8	**1.7**



#31 #63 #90 #226 #534 #622

(a)

#58 #154 #266 #428 #473 #716

(b)

#3 #171 #208 #233 #240 #294

(c)

#10 #65 #172 #280 #353 #374

(d)

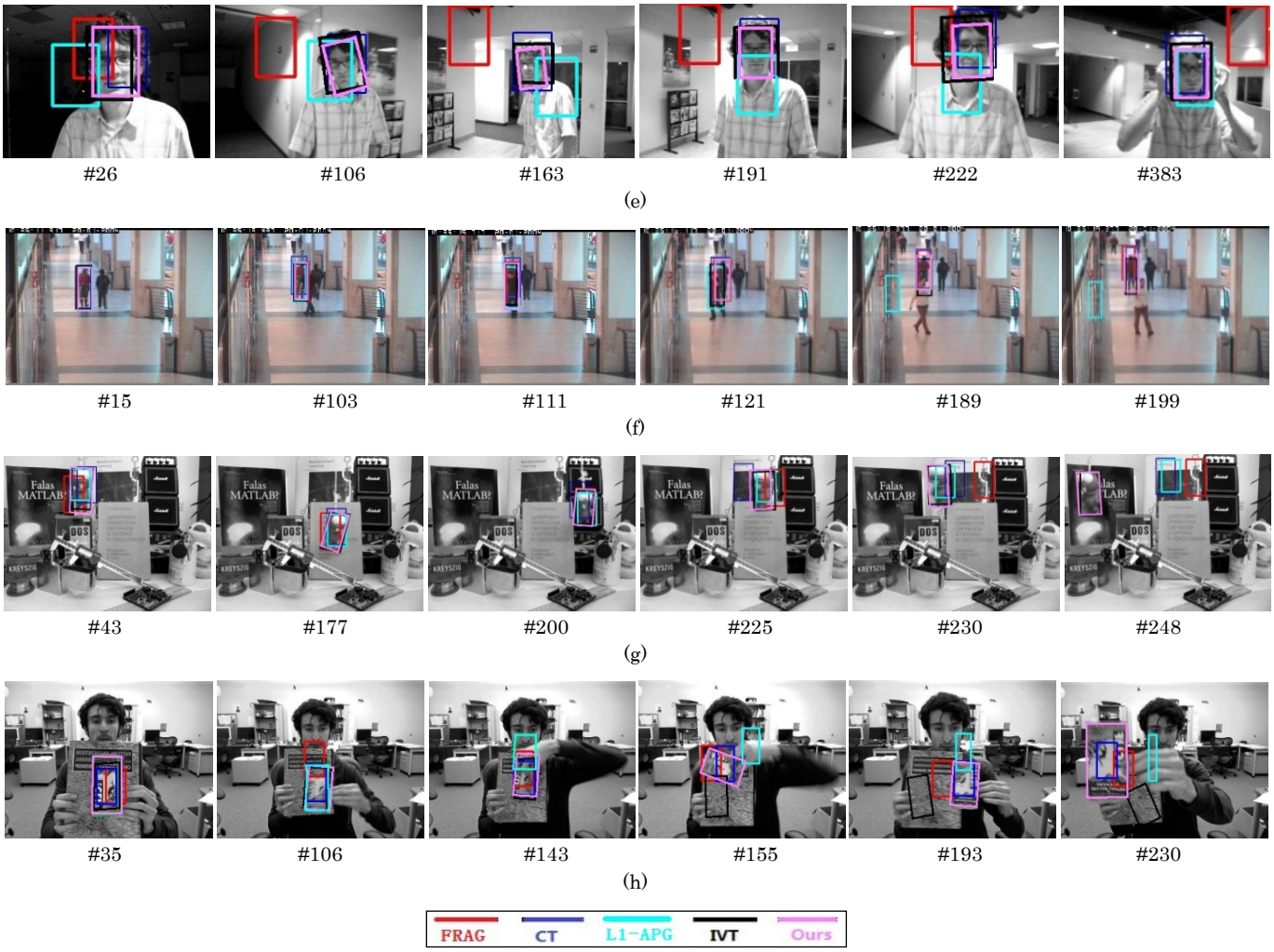


Fig. 4 Tracking results of different algorithms for sequences (a) *Occlusion1* , (b) *Occlusion2*, (c)*car4*, (d)*car11*, (e) *David_indoor*, (f) *Caviar1*, (g) *Lemming* , and (h) *Cliffbar*.

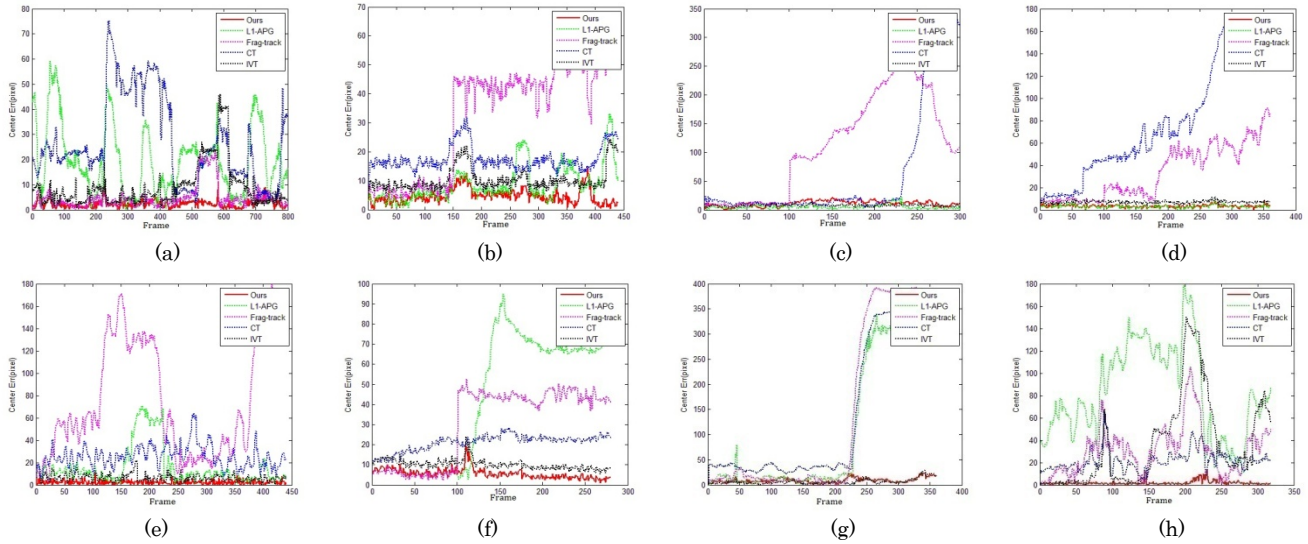


Fig. 5 Quantitative evaluation of different trackers in terms of center location error on sequences (a) *Occlusion1* , (b) *Occlusion2*, (c)*car4*, (d)*car11*, (e) *David_indoor*, (f) *Caviar1*, (g) *Lemming* , and (h) *Cliffbar*.

6.Conclusion

In this paper, we employ Bi-2DPCA for appearance modeling in a tracking framework. This scheme allows us to take advantage of spatial information of the raw intensity image and avoid calculating the large size covariance matrix. We proposed a new incremental Bi-2DPCA subspace learning technique which models the appearance changes of an object. Furthermore, we introduced a sparse structure into our Bi-2DPCA based subspace to handle noise and occlusions. The time-consuming l_1 regularized optimization procedure is replaced by an iterative algorithm to achieve a real-time implementation. Experimental results on the challenging image sequences demonstrate that our tracking algorithm performs favorably against state-of-the –art algorithms. Our future work will focus on utilizing prior knowledge with online learning for more effective object tracking.

This work was supported by the Research Fund for the Key Project of Technology Research Plan of Ministry of Public Security, China under Grant 2014JSYJA018, the Natural Science Research Project of Education Department of Shaanxi Province, China under Grant 12JK0731, the Research Fund for the Doctoral Program of Higher Education of China under Grant 20126102110041, and the Royal Academy of Engineering, UK, under Grant 1314RECI025.

References

1. D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental learning for robust visual tracking," Int. J. Comput. Vision, **77**, 125–141 (2008).
2. D. Wang, H. Lu, and Y.-W. Chen, "Incremental MPCA for color object tracking," in *Proceedings of IEEE Conference on Pattern Recognition* (IEEE, 2010), pp. 1751–1754.
3. W. Hu, X. Li, X. Zhang, X. Shi, S. J. Maybank, and Z. Zhang, "Incremental tensor subspace learning and its applications to foreground segmentation and tracking" Int. J. Comput. Vision, **91**, 303–327 (2011).
4. D. Ross, J. Lim, and M. Yang, "Adaptive probabilistic visual tracking with incremental subspace update," in *Proceedings of the Eighth European Conference on Computer Vision* (IEEE, 2004), pp. 470–482.
5. G. Li, D. Liang, Q. Huang, S. Jiang, and W. Gao, "Object tracking using incremental 2D-LDA learning and Bayes inference," in *Proceedings of IEEE Conference on Image Processing* (IEEE, 2008), pp. 1568–1571.
6. T. Can, A. O. Karal, and T. Aytac, "Detection and tracking of sea-surface targets infrared and visual band videos using the bags-of-features technique with scale-invariant feature transform, "Appl. Opt. 50, 6302–6312 (2011).
7. S. Avidan, "Ensemble tracking," IEEE Trans. Pattern Anal. Mach. Intell., **29**, 261–271 (2007).
8. Y. Li, L. Xu, J. Morphett and R. Jacobs, "On Incremental and Robust Subspace Learning," Pattern Recognition, **37**, 1509-1518 (2004).
9. D. Skocaj, A. Leonardis, "Weighted and Robust Incremental Method for Subspace Learning," in *Proceedings of IEEE Conference on Computer Vision* (IEEE, 2003), pp.1494-1501.
10. J. Yang, D. Zhang. "Two-Dimensional PCA: A new approach to appearance-based face representation and recognition," IEEE Trans. Pattern Anal. Machine Intell. **26**, 131-137 (2004).
11. T. Wang, I. Gu, and P. Shi, "Object tracking using incremental 2D-PCA learning and ML estimation," in *Proceedings of IEEE Conference on Acoustics Speech Signal Processing* (IEEE, 2007), pp. 933–936.
12. Q. Wang, F. Chen, and W. Xu, "Visual tracking by appearance modeling and sparse representation," in *Proceedings of IEEE Conference on Neural Computation* (IEEE, 2010), 1464-1468.
13. D. Zhang, Z.-H. Zhou, "$(2D)^2$ PCA: Two-directional two-dimensional PCA for efficient face representation and recognition, " Neurocomputing, **69**, 224-231(2005).
14. J. Yang, Y. Xu and J. Yang, "Bi-2DPCA: a fast face coding method for recognition", Pattern Recognition Recent Advances, pp. 313-340, Adam Herout (Ed.), InTech (2010).
15. D. Wang, H. Lu, and M. Yang, "Online Object Tracking with Sparse Prototypes," IEEE Trans. Image Process. **22**, 314-325 (2013).
16. X. Mei and H. Ling, "Robust Visual Tracking using L_1 minimization," in *Proceedings of IEEE Conference on Computer Vision* (IEEE, 2009), pp. 1436–1443.
17. Y. Li, P. Li, and Q. Shen, "Real-time infrared target tracking based on l_1 minimization and compressive features," Appl. Opt. **53**, 6518–6526 (2014).
18. X. Jia, H. Lu, M Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proceeding of IEEE Conference on Computer Vision and pattern Recognition* (IEEE, 2012), pp. 1822-1829.
19. X. Mei, and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," IEEE Trans. Pattern Anal. Machine Intell., **33**, 2259-2272 (2011).
20. W. Zhang, H. Lu, and M. Yang, "Robust object Tracking via sparse collaborative appearance model," IEEE Trans. Image Process. **23**, 2356-2368 (2014).
21. X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai, "Minimum error bounded efficient l_1 tracker with occlusion detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 1257-1264.
22. C. Bao, Y. Wu, H. Ling, and H. Ji, "Real-time robust l_1 tracker using accelerated proximal gradient approach," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012), pp. 1830-1837.
23. Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Online Object Tracking: A Benchmark," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2013) pp, 2411 - 2418.
24. A. Adam, E. Rivlin, and I. Shimshoni. "Robust Fragments based Tracking using the Integral Histogram," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2006) pp, 798–805.
25. K. Zhang, L. Zhang, and M. Yang, "Real-Time Compressive Tracking," in *Proceedings of European Conference on Computer Vision* (2012), pp.864-877.